

Regulations, and Responsibility, and Respect:

On the Ethical Adequacy of the T3 Study

Erin M. Ochoa

MACS 30000: Perspectives on Computational Analysis

2016 October 10

Research ethics are a vital part of any study and are especially salient for those involving human subjects. Scientists have a responsibility to ensure that the studies they design and the data-collection methods they implement conform to established ethical standards. In *Bit By Bit*, Matthew J. Salganik¹ describes four such principles for ethical research: Respect for persons, beneficence, justice, and respect for law and public interest. Here we evaluate the extent to which the Harvard “Tastes, Ties, and Time” (henceforth *T3*) study² adhered to each of these principles.

Assessing conformity to the first principle, respect for persons, is straightforward: Considering that the source of the data was correctly identified as Harvard College mere days after the release of the codebook,³ we find that researchers’ failure to effectively protect subjects’ privacy is tantamount to disrespect. The researchers knew or should have known that the nature of some of the data, such as uniquely named fields of study,⁴ would quickly lead to re-identification of the data set and therefore a breach of subjects’ privacy. In their comments (see Zimmer⁵ for examples), however, it is apparent that researchers had both an inflated view of the inscrutability of the data set’s origin and a shortsighted appraisal of potential harms that could befall subjects should re-identification occur. While we appreciate that researchers went to what they considered to be

¹ Matthew J. Salganik, *Bit By Bit: Social Research in the Digital Age* (<http://www.bitbybitbook.com>, in review), 6.4.

² Kevin Lewis, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, and Nicholas Christakis, “Tastes, Ties, and Time: A New Social Network Dataset using Facebook.com” in *Social Networks* (<http://dx.doi.org/10.1016/j.socnet.2008.07.002>, 2008).

³ Michael Zimmer, “‘But the Data is Already Public’: On the Ethics of Research in Facebook” in *Ethics and Information Technology* (<http://dx.doi.org/10.1007/s10676-010-9227-5>, 2010), 316.

⁴ Ibid., 316–317.

⁵ Ibid.

great lengths to anonymize the data, we find that outcomes matter more than intentions. Furthermore, that one researcher acknowledged the team's lack of technical expertise⁶ is especially telling and raises the question of why they failed to consult with technical experts about effective anonymization methods or the possible consequences of re-identification.⁷ Had researchers felt respect for their subjects as real persons, perhaps they would have more thoroughly evaluated the potential risks or even asked subjects' consent before collecting data. Though this may have diminished the size of the data set—and, potentially, its usefulness—such a compromise would have been warranted, given the sensitive nature of the data.

The second principle of ethical research, beneficence—perhaps the most complex of the four—has three main tenets: do not harm, ensure that benefits outweigh possible risks, and apply for institutional review to ensure a proper balance between benefits and risks.⁸ With regard to the first tenet, the T3 research team failed: Because subjects' privacy was inadequately protected, subjects were made vulnerable to a wide range of potential harms. While enumerating such possible harms or determining whether any occurred is beyond the scope of this review, we believe that rendering subjects susceptible to such risks is enough to find that the tenet went unfulfilled.

Similarly, the T3 study did not adhere to the ethical requirement of maximizing benefits, minimizing possible risks, and ensuring that benefits outweighed risks, because researchers did not identify a clear and meaningful purpose that balanced the potential risks. In the absence of a compelling research reason, data of such a sensitive nature should not have been collected. While we recognize that the original intended purpose for a data set may not necessarily be its most

⁶ Ibid., 316.

⁷ Ibid., 321.

⁸ Salganik, *Bit By Bit: Social Research in the Digital Age* (in review), 6.4.

important one, this does not mean that we should collect data haphazardly simply because we can. Instead, researchers must begin with a compelling reason to collect the data in the first place, and the strength of that reason must increase as the degree of informed consent is diminished and the privacy risks rise: If subject consent is not required, then the reason for collecting data must be critical; teasing out the relationship between social networks and members' tastes in books, movies, and music is surely insufficient. Additionally, if the potential consequences of re-identification are high (as Zimmer⁹ proposes they were for this study), then the potential benefits to subjects must be abundant and significant in order to balance those risks.

The third tenet of beneficence, which requires that researchers apply for institutional review to ensure that potential risks and benefits are properly balanced, is not easy to evaluate because to do so thoroughly would require access to documents that are not publicly available. We do know that the T3 research team did seek and obtain approval from the Harvard College institutional review board, the Committee on the Use of Human Subjects¹⁰ (henceforth *IRB*). Without reviewing the *IRB* application, however, we cannot ascertain whether the research team explicitly disclosed that the research assistants who would collect the data were potentially members of some subjects' social networks and therefore had access to data not otherwise publicly available.¹¹ If researchers did raise this, then they at least attempted to identify an important privacy concern; if they did not mention this in the application, then they clearly and inexcusably engaged in collecting data without understanding the workings of the platform they were accessing. Because we cannot come to a determination, we must find neither adherence to nor departure from this tenet. When we consider

⁹ Zimmer, "But the Data is Already Public": On the Ethics of Research in Facebook", 319.

¹⁰ *Ibid.*, 315.

¹¹ *Ibid.*, 318.

that two of the three tenets of beneficence went unfulfilled, though, we must find that the principle was not met.

Salganik's justice principle requires that the burdens and benefits of research be distributed equitably.¹² The T3 study fails to meet this principle because the burdens of research were imposed solely upon the students entering Harvard College in 2006. The collection methods likely resulted in the inclusion of information that would not have otherwise been accessible. Furthermore, the individual circumstances of students who could be more easily identified (such as international students and students with no home-state peers¹³) left them at an increased risk. Again, we do not find a compelling interest, either for the public at large or the subjects of the data set, to justify collecting the data at all. While the data collected for T3 are likely invaluable for product marketing teams or marketing researchers, potential benefits for the public at large or the research subjects are not clear. Because the research team failed to identify a pressing concern about subjects' lives that could be addressed by the research, the study falls short of meeting the spirit of the justice principle.

The T3 study partially adheres to the principle of respect for law and public interest. In obtaining consent from the owner of the data—Facebook—and from the Harvard IRB, the team was likely in the clear, at least legally. Establishing respect for public interest, however, also requires the research team to exercise “transparency-based accountability”.¹⁴ As evinced by researchers' comments (see Zimmer¹⁵ for examples) regarding the so-called extreme difficulty of re-identification, they were focused on the potential harms of such re-identification but either did

¹² Salganik, *Bit By Bit: Social Research in the Digital Age* (in review), 6.4.

¹³ Zimmer, “‘But the Data is Already Public’: On the Ethics of Research in Facebook”, 319.

¹⁴ Ibid.

¹⁵ Zimmer, “‘But the Data is Already Public’: On the Ethics of Research in Facebook”.

not conceive of or did not acknowledge the inherent dignity concerns of accessing private information without gaining subjects' explicit and informed consent.¹⁶

Having reviewed the extent to which the T3 study met or failed to meet each of Salganik's principles for ethical research, we conclude that the study, as conducted, did not constitute ethical research. Given that conclusion, we would decline to use the T3 data for research. To do otherwise would perpetuate a violation of subjects' privacy without their consent and, potentially, cause harm. Furthermore, it is unlikely that any institutional review board would approve the use of the data set, especially considering the controversy in which it is mired.

¹⁶ Ibid., 321.

Bibliography

- Lewis, Kevin, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, and Nicholas Christakis. “Tastes, Ties, and Time: A New Social Network Dataset using Facebook.com.” *Social Networks* 30, 4 (October 2008): 330–342. <http://dx.doi.org/10.1016/j.socnet.2008.07.002>.
- Salganik, Matthew J. *Bit By Bit: Social Research in the Digital Age*. <http://www.bitbybitbook.com>. In review.
- Zimmer, Michael. “‘But the Data is Already Public’: On the Ethics of Research in Facebook.” *Ethics and Information Technology* 12, 4 (December 2010): 313–325. <http://dx.doi.org/10.1007/s10676-010-9227-5>.